

Information measures and cognitive limits in multilayer navigation

Riccardo Gallotti¹, Mason A. Porter^{2,3}, and Marc Barthelemy^{1,4*}

¹Institut de Physique Théorique, CEA-Saclay, F-91191, Gif-sur-Yvette, France

²Oxford Centre for Industrial and Applied Mathematics,

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

³CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK

⁴Centre d'Analyse et de Mathématique Sociales, EHESS,

190-198, Avenue de France, 75244 Paris Cedex 13, France

*To whom correspondence should be addressed; E-mail: marc.barthelemy@cea.fr.

Cities and their transportation systems become increasingly complex and multimodal as they grow, and it is natural to wonder if it is possible to quantitatively characterize our difficulty to navigate in them and whether such navigation exceeds our cognitive limits. A transition between different searching strategies for navigating in metropolitan maps has been observed for large, complex metropolitan networks. This evidence suggests the existence of another limit associated to the cognitive overload and caused by large amounts of information to process. In this light, we analyzed the world's 15 largest metropolitan networks and estimated the information limit for determining a trip in a transportation system to be on the order of 8 bits. Similar to the “Dunbar number,” which represents a limit to the size of an individual's friendship circle, our cognitive limit suggests that maps should not consist of more than about 250 connections points to be easily readable. We also show that including connections with other transportation modes dramatically increases the information needed to navigate in multilayer transportation networks: in large cities such as New York, Paris, and Tokyo, more than 80% of trips are above the 8-bit limit. Multimodal transportation systems in large cities have thus already exceeded human cognitive limits and consequently the traditional view of navigation in cities has to be revised substantially.

Introduction

The number of “megacities”—urban areas whose human population is larger than 10 million—has tripled since 1990 (1). New York City, one of the first megacities, reached that level in the 1950s, and the world now has almost 30 megacities, which together include roughly half a billion inhabitants. The growth of such large urban areas usually also includes the development of transportation infrastructure and an increase in the number and the use of different transportation modes (2). For example, about 80% of cities with populations larger than 5 millions have a subway system (3). This leads to a natural question: Is navigating transportation systems in very large cities too difficult for humans (4)? Additionally, how does one quantitatively characterize this difficulty?

It has long been recognized that humans have intrinsic cognitive limits for processing information (5). In particular, it has been suggested that an individual can maintain only on the order of 150 stable relationships (6, 7). This “Dunbar number” was first proposed in the 1990s by British anthropologist Robin Dunbar by extrapolating results about the correlation of brain sizes and the typical sizes of social groups for various primates. Although it is still controversial, it has been supported by subsequent studies on, e.g., traditional human societies (8) and microblogging (9).

When navigating for the first time between two unfamiliar places and having a transportation map as one’s only support, a traveler has to compare different path options to find an optimal route. Differently from the case of the schematization of partially familiar routes (10), here the traveler does not need to simultaneously visualize the whole route; it is sufficient to identify and keep track of the position of the connecting stations on the map. Therefore, a first important point to consider is that humans can track information on a maximum of about four objects in their visual working memory (11). This implies that a person can easily keep in mind the key locations (origin, destination, and connection points) for trips with no more than two connections (which corresponds exactly to four different points). In addition, recent studies on visual search strategies (12, 13) show a transition in search strategies between the simple cases of the Stuttgart and Hong Kong metropolitan networks and the case of Paris, which has one of the most complicated transportation networks in the world. The time needed to find a route in a transportation network grows with the complexity of its map, and the pattern of eye fixations also changes from following metro lines to a random scattering of eye focus all over the map (12). A similar transition from directional to isotropic random search has been observed for visual search of hidden objects when increasing the number of distractors (14). The ability to manage complex “mental maps” is thus limited, and only extensive training on spatial navigation can push this limit with morphological changes in the Hippocampus (15). Human-constructed environments have far exceeded these limits, and it is interesting to ask whether there is a navigation analog of the Dunbar number and a cognitive limit to human navigation ability, such that it becomes necessary to rely on artificial systems to navigate in transportation systems in large cities. If such a cognitive limit exists, what is it? In this paper, we answer this question using an information perspective (5) to characterize the difficulty to navigate in urban transportation

networks. We use a measure of “information search” associated to a trip that goes from one route to another (16). In most networks, many different paths connect a pair of nodes, and one generally seeks the fastest path that minimizes the total time to reach a destination. However, it tends to be more natural for most individuals to instead consider a “simplest” path, which has the minimum number of connections (18) (see Fig. 1).

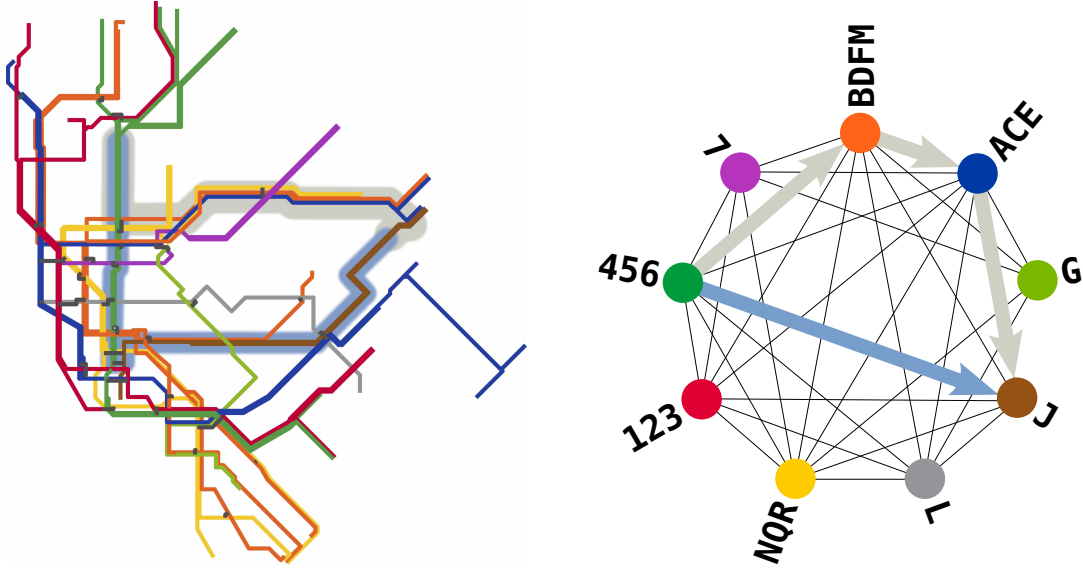


Figure 1: **Fastest and Simplest paths in Primal and Dual networks.** (Left) In the primal network of the New York City metropolitan system, the simplest path (highlighted in light blue) from 125th St. on line 5 (dark green) to 121st St. on line J (brown) differs significantly from the fastest path (highlighted in grey). There is only one connection for the simplest path (*Brooklyn Bridge – City Hall / Chambers Street*) in Lower Manhattan. In contrast, the fastest path needs three connections (5 → F → E → J). We compute the fastest path using travel times from the Metropolitan Transportation Authority (MTA) Data Feeds (see Materials and Methods). We neglect walking and waiting times. (Right) In the dual space, nodes represent routes and edges represent connections. A simplest path in the primal space is defined as a shortest path with the minimal number of edges in the dual space (light blue arrow). It has a length of $C = 1$ and occurs along the direct connection between line 5 (dark green node) and line J (brown node). The fastest path in the primal space has a length of $C = 3$ (grey arrows) in the dual space, as one has to change lines three times. (We extracted the NYC metro schematic from a map that is publicly available on Wikimedia Commons (35).)

Rosvall *et al.* (16) proposed a measure for the information that is needed to encode a shortest path from a route s to another route t . However, the amount of necessary information can depend strongly on the initial and final nodes, and we consider a trip from an origin node i in route s to a destination node j in route t . This trip is embedded in real space and among all possible simplest paths (16) (which need not to be unique), we pick the fastest one $p(i, s; j, t)$,

which can differ from an actual fastest path between i and j (see the left panel of Fig. 1). For computing the travel time of a trip, we neglect the contribution of walking and waiting times (2). However, the choice of a simplest path already tends towards minimization of such transfer costs, which strongly influence a traveler’s decisions (17).

The total information for knowing the fastest simplest path is

$$S(i, s; j, t) = -\log_2 \left(\frac{1}{k_s} \prod_{n \in p(i, s; j, t)} \frac{1}{k_n - 1} \right), \quad (1)$$

where $p(i, s; j, t)$ is the sequence of routes needed for connecting i in route s to j in route t . The term k_s is the number of routes connected to s , and along the path, we have the choice between $k_n - 1$ routes. The idea behind Eq. (1) is that when tracking a trip along a map (with the eyes or a finger), the connections that one has to exclude represent—similarly to the number of distractors in visual search tasks (14)—the information that has to be processed and thus temporarily stored into working memory (19). One can therefore construct the measure of entropy (1) as a proxy for the accumulated cognitive load that is associated to the trip, and it is analogous to the total amount of load experienced during a task (20). For this reason, this measure of entropy seems to be appropriate for estimating the information limit associated to the observed transition in the visual search strategy (12, 14).

From a map user’s perspective, the existence of several alternative simplest paths is not necessarily a significant factor, as one only needs a single simplest path for successful transportation from origin to destination. Consequently, we use the entropy in Eq. (1) rather than the one proposed in Ref. (16). (See the Materials and Methods for further discussion.) To produce a single summary statistic for a path, we average $S(i, s; j, t)$ over all nodes $i \in s$ and $j \in t$ (we denote this mean using brackets $\langle \cdot \rangle$) to obtain

$$\bar{S}(s, t) = \langle S(i, s; j, t) \rangle, \quad (2)$$

which is the main quantity that we use to describe the complexity of a trip (see Fig. 5) and which will allow us to extract an empirical upper limit to the information that a human is able to process for navigating.

Results

We use the measure $\bar{S}(s, t)$ to characterize the complexity of the 15 largest urban metropolitan systems in the world. From these results, we then extract an empirical upper limit of about 8 bits to the information that a human is able to process for navigating. We then apply this threshold in calculations for multimodal transportation networks and demonstrate that most trips in large cities exceed human cognitive limits.

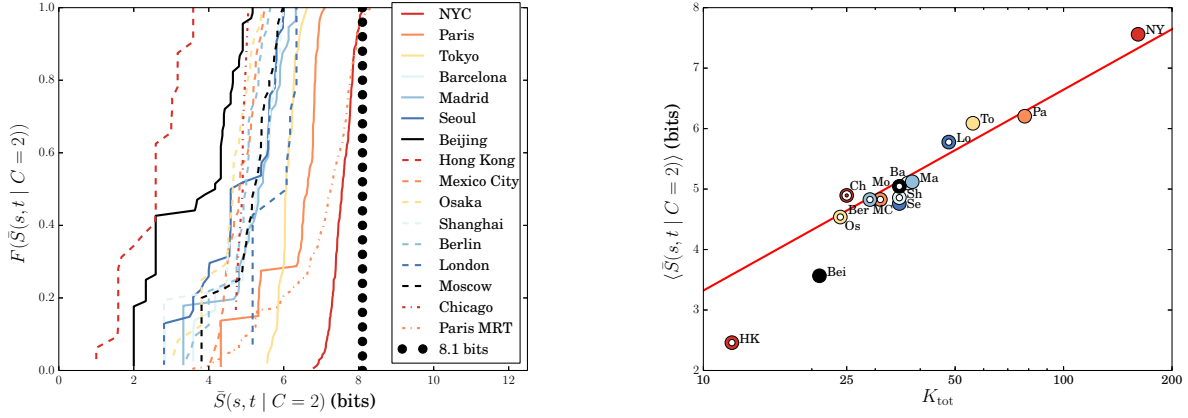


Figure 2: **(Left) Information threshold.** Cumulative distribution of the information needed to encode trips with two connections in the 15 largest metro networks. The largest value occurs for the New York City metro system (solid red curve), which has trips with a maximum of $S_{\max} \approx 8.1$ bits. Among the 15 networks, the Hong Kong (dashed, red) and Beijing (solid, black) metro networks have the smallest number of total connections and need the smallest amount of information for navigation. The Paris MRT (Metro, Light Rail, and Tramway) network (orange, dash-dotted) from the official metro map (32), includes three transportation modes (which are managed by two different companies) and reaches values that are similar to those in the larger NYC Metro. **(Right) Information threshold versus total number of connections in the dual space.** This plot illustrates that the mean amount of information that is needed to encode trips with two connections is strongly correlated with the total number of connections in the dual network, as can be predicted for a square lattice (see Materials and Methods). See Table 1 for the definitions of the abbreviations. The color code is the same as in the left panel, and the red solid line represents the square-lattice result $\bar{S}(s, t | C = 2) = \log_2(K_{\text{tot}})$. This relationship permits one to associate the information threshold S_{\max} with the cognitive threshold $T = 2^{S_{\max}}$, which one can interpret as a maximum number T of intra-route connections that can be represented on a map.

Information Threshold

The values of $\bar{S}(s, t)$ in a network tend to grow with the number C of connections that appear in a simplest path as well as with the mean degree $\langle k \rangle$ of the nodes in the dual space (see Figs. 6 and 7). Note that the latter is related to the total number of connections in a network. Adding new routes can thus have a negative impact from the information perspective. Although new routes can be useful for shortening the simplest paths for some (s, t) pairs, new connections simultaneously increase the mean degree of a network and can make it more difficult to navigate in a network. We thus want to estimate the maximum possible information that an individual can reasonably process to navigate in a transportation system. For that purpose, we consider

the world's 15 metro networks with the largest number of stations. The characteristics of metro networks were examined in previous papers (3, 21–26), and navigation strategies have been considered in transportation networks (27–29). For each network, we consider the shortest simplest paths with $C = 2$ connections. This corresponds to paths that use 3 different lines: such a path starts from a source route s , connects to an intermediate route r , and then connects to a destination route t . There are two distinct reasons for this choice: (i) the limit of four objects in the visual working memory (11); (ii) in most of the 15 cities, two connections correspond to the diameter of the dual network. From a map user's perspective, after having checked that each pair of consecutive stations are connected by a direct line, the locations to keep in mind are the origin, the destination, and connecting stations. These nodes correspond to the places that one has to “highlight” on the map to record the trajectory. The capacity of visual working memory thus allows one to easily keep in mind only trajectories with 2 connections, leading to a total of 4 stations. Interestingly, the value 2 is also the diameter of the dual network in most of the metropolitan systems. In such situations, all pair of nodes can be connected by paths with at most 2 connections, staying below the working memory limit of humans.

In the left panel of Fig. 2, we show the cumulative distribution of entropies $\bar{S}(s, t|C = 2)$ for these 2-connections paths. We find that the New York City metro system is the largest and most complex metropolitan system in the world; it has a maximal value of $S_{\max} \approx 8.1 \approx \log_2(274)$ bits. Paris' transportation system reaches a similar value if one takes into account the light rail and tram system in the multilayer Metro-Rail-Tramway (MRT) network displayed in the official metro map (32).

Navigation in such large networks is already nontrivial (13), and it has been observed that there is an eye-movement behavioral transition when the system becomes too large (i.e., when there are too many connections) (12). The value S_{\max} for trips with two connections thus provides a natural limit, above which human cognitive capabilities are challenged and for which it becomes extremely difficult to find a simplest path. We thus make the reasonable choice to take S_{\max} as the cognitive limit for public transportation: a human needs an information entropy of $\bar{S}(s, t) \leq S_{\max}$ to be able to navigate in a network successfully without assistance from information technology tools.

To gain a physical understanding for the cognitive limit S_{\max} , we estimate $S(s, t|C = 2)$ for a regular lattice (like the one in Fig. 3) with N lines that are each connected with $N/2$ other lines (i.e., $k_r = N/2$ for all r). This choice of a lattice is justified by the results in Ref. (3) that most large metropolitan transportation networks consist of a core set of nodes with branches that radiate from it. The core is rather dense and has a peaked degree distribution, so it is reasonable to use a regular lattice for comparison. In the dual space of the regular lattice, the degree k_s of route s is equal to $N/2$, and we thus obtain

$$\begin{aligned} \bar{S}(s, t|C = 2) &= \log_2[k_s(k_r - 1)] \\ &\approx \log_2(\langle k \rangle^2) = \log_2 \left(\sum_{i=1}^N k_i/2 \right), \end{aligned} \quad (3)$$

where $\langle k \rangle$ denotes the mean degree. The last equality in Eq. (3) comes from the relation for the total degree of a regular lattice: $\sum_{i=1}^N k_i = \langle k \rangle N = 2\langle k \rangle^2$. The key quantity for understanding S_{\max} is therefore the total number of undirected connections $K_{\text{tot}} = \sum_{i=1}^N k_i/2$ in the dual space (also see Table 1). As we indicated in Eq. (3), this is identical to the square of the mean degree $\langle k \rangle^2$ in a lattice. For Paris, for example, we obtain $\langle k \rangle \approx 9.75$, which leads to $9.75^2 \approx 95$ connections for the corresponding lattice. The actual Paris metropolitan network has a total of 78 connections, and the difference comes from the fact that the real network is not a perfectly regular lattice. At this stage, it is important to make two remarks. First, the apparently paradoxical fact that the total information grows with size for regular lattices while intuitively the complexity for finding a path stays constant is specific to the case in which there is an “algorithm” to find the route. Indeed, in perfectly regular rectangular (i.e., Cartesian) grids, one needs to make at most two turns to find a desired route. Second, drawing a parallel with an individual navigating a bifurcating tree from the root to a terminal node who is traversing 8 bifurcation points (and thus 255 internal nodes), the value of 8 bits as a limit appears to be consistent with Miller’s “magic number” (5,30). “Miller’s law” limits also the number of binary decisions that can be memorized in sequence to a value that has been observed to be in the range 7 ± 2 .

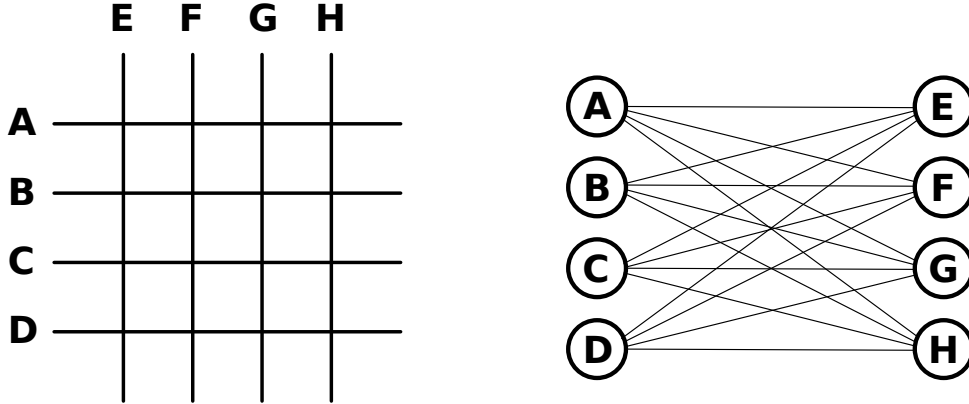


Figure 3: **(Left) Primal and (right) dual networks for a square lattice.** In this example, the lattice has $N = 8$ routes. Each route has $k = N/2 = 4$ connections, so the total number of connections is $K_{\text{tot}} = k^2 = 16$. In the dual network, the four East-West routes (A,B,C,D) and the four North-South routes (E,F,G,H) yield a graph with a diameter of 2.

We test Eq. (3) in the right panel of Fig. 2 for the 15 largest metropolitan networks. Our calculation shows that the total degree in the dual space, which is related to the total number of connections in a network, is the main ingredient for understanding the information entropy of these systems. We also test this relation for the temporal evolution of the Paris metro network and we show that the number of connections $\sum_{i=1}^N k_i/2$ scales as $(N/2)^2$ for the historical growth from $N = 1$ to $N = 14$ routes (see Fig. 8). Equation (3) allows us to translate the information limit of 8 bits to a limit on the number T of intra-route connections ($S = \log_2(T)$).

The value of T also corresponds to the number of distractors to be excluded for the most complex trips (with $C(s, t) = 2$). This process of exclusion thus demands progressive information integration, which causes a cognitive overload.

Evidence for the existence of such a cognitive threshold is the change in search strategy observed in eye-tracking experiments (12, 14). The value $T \approx 250$ represents the worst-case scenario in the world’s largest metropolitan network. It thus overestimates the values at which the transition occurs. Indeed, the Paris MRT network, for which the strategy change was observed in (12), has $K_{\text{tot}} \approx 162$. It is interesting that T has a similar order of magnitude as the Dunbar number, an extensively studied cognitive limit for the size of a friendship’s circle, and which seems to lie in a range between 100 and 200. (See, for example, (9) for a recent discussion of this topic.).

Effect of Multimodal Couplings on the Information

In our discussion above, we estimated the cognitive threshold for the most complex paths in the 15 largest metropolitan networks. We now consider the effect of including other transportation modes (e.g., buses, trams, etc.). The effects of inter-modal coupling are significant (2, 31), and the natural framework is a multilayer network (33, 34), which associates each transportation mode with a different “layer” in a network and where interchanges (i.e., connection points) between different modes are represented by inter-layer edges. As we discuss in the Materials and Methods section, a major difficulty comes from obtaining data for different modes for a given city, as this type of data is not currently available in general. However, we were able to combine multiple data sources for large cities on three different continents: New York City, Paris, and Tokyo (see Table 2).

The distribution of $\bar{S}(s, t)$ is a superposition of peaks associated to different values of C (see Fig. 6). By comparing the distribution of $\bar{S}(s, t)$ for the bus monolayer network and the full multilayer transportation network, we can distinguish two competing effects of multimodality: (i) it tends to reduce the number C of connections and thereby reduces \bar{S} ; and (ii) it increases \bar{S} , because new routes increase the node degrees in the dual space. (See Fig. 7) However, these two contributions do not compensate each other. In Fig. 4, we show the cumulative distribution $F(\bar{S}(s, t))$ of information entropy values for the New York, Paris, and Tokyo multimodal (i.e., multilayer) transportation systems.

We find that less than approximately 17% (the maximum value obtained for Tokyo) of the trips are below the threshold S_{max} . These results imply that more than 80% of the trajectories in the complete public transportation networks of these major cities require more information than the most complicated trajectory in the largest metro networks. As we show in the Supplementary Materials, the other 20% correspond to pairs of nodes for which the trip has essentially one connection (for NYC) or at most two connections (for Paris and Tokyo), as the simplest paths that carry a small amount of information are those that avoid using too many major hubs. (See Table 4 and Fig. 9). The number of connections acting as distractor for the case of the Paris MRT is already so large that it has a crucial impact in the route search (it takes roughly 30

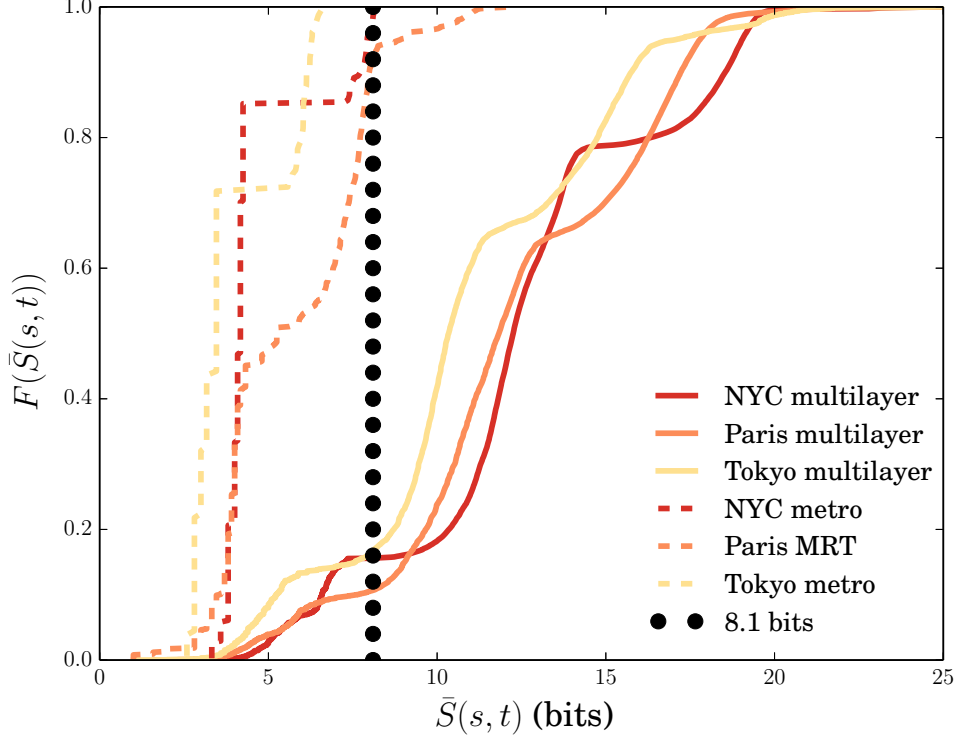


Figure 4: **Information entropy of multilayer networks.** The solid curves represent the cumulative distributions of $\bar{S}(s, t)$ for multilayer networks that include a metro layer for New York City, Paris, and Tokyo. (We associate one layer to bus routes and another to metro lines.) Most of the trips require more information than the cognitive limit $S_{\max} \approx 8.1$. The fraction of trips under this threshold are 15.6% for NYC, 10.7% for Paris, and 16.6% for Tokyo. The dashed curves are associated to all possible paths in a metro layer; in this case, the amount of information is always under the threshold, except for Paris, which includes trips with $C = 3$. (See Table 3). Note that the threshold value lies in a relatively stable part of all three cumulative distributions, which suggests that our results are robust with respect to small variations of the threshold.

seconds on average for such as search (13)), and the complexity of the bus layer (and therefore of the coupled metro and bus system) will therefore exceed the human capacity. Consequently, traditional maps that represent all existing bus routes have a very limited utility. This result thus calls for the need of thinking about a user-friendly way to present and to use bus routes. For example, unwiring some bus-bus connections lowers the information, and leads to the idea that a design centered around the metro layer could be efficient. Further work is however needed to reach an efficient, 'optimal' design from a user perspective.

Finally, we note that one can also think of finding an optimal path by separately considering

different transportation modes. The price to pay for such a strategy would be to settle for suboptimal trajectories (which are neither simplest nor shortest), as an optimal trajectory is, by definition, computed on a full multilayer transportation network.

Discussion

Human cognitive capacity is limited, and cities and their transportation network have grown to the point that they have reached a level of complexity that is beyond human processing capability to navigate in them. In particular, the search for a simplest path becomes inefficient when multimodality is important and when a transportation system has too many interconnections. This occurs because of interconnections, which play two crucial roles in the search for a path: they are both targets and distractors. The identification of possible interchange points is a key and extremely time-consuming passage in the route-finding task (12). As with the case of hidden objects (14), one can represent the difficulty of the search using the number of distractors, which for a map are all possible interchanges. We have found in the largest cities that the addition of bus routes with maps that are already too complicated to be easily used by travelers implies that the cognitive limit to urban navigation is exceeded for multimodal transportation systems.

We have estimated the cognitive threshold value to be of $T \approx 250$ connections (i.e., approximately 8 bits), which represents the worst-case scenario for the most complex trips in the largest networks. This exceeds the behavioral transition, beyond which humans have difficulties to navigate on their own.

Remarkably, we note that T has a similar interpretation and order of magnitude as the Dunbar number, and our result can thus be construed as evidence to suggest the existence of a “transportation Dunbar number” for processing complex maps. Although the value of T represents an upper bound, it allows us to demonstrate the existence of the huge gap between the amount of information that a human can process easily ($K_{\text{tot}} < 250$) and the amount that is contained in practice in multimodal transportation networks in large cities ($K_{\text{tot}} > 1800$, see Table 2). Indeed, the growth of transportation systems has yielded networks that are so entangled with each other and so complicated that a visual representation on a map becomes too complex and ultimately useless. In particular, our results imply that maximizing the number of intersections between lines, which indeed minimizes transfers, is contrary to the goal of having an easily usable transportation system. The information-technology tools provided by companies and transportation agencies that help people to navigate in transportation systems will soon become necessary in all large cities. Our analysis highlights the fact that humans need to integrate navigation complexity and seek new solutions that will help for navigation in megacities. Redesigning maps and representations of transportation networks (36) as well as improving information-technology tools (37) that help to decrease the amount information below the human processing threshold thus appear to be crucial for an efficient use of services provided by transportation agencies.

Materials and Methods

Data for the World’s 15 Largest Metropolitan Transportation Networks

The data that we use was extracted from Wikipedia (38). It describes the lines in the metropolitan networks as they were in 2009, and the data were used in a previous publication (3). The data for each metropolitan system yields a spatial network (39); each station has geographical coordinates, and the edges connect consecutive stations on a metro line. The diagnostic that we use for determining a shortest simplest path is the sum of the traveling times between stops.

Paris

The data that we use for constructing the Paris multilayer network (33, 34) come from the Régie Autonome des Transports Parisiens (40) and the Société Nationale des Chemins de fer Français (41). Both data sets are provided in GTFS format. We extract the bus layer from the RATP data. To reproduce all of the information available in the Paris Metro Map (32), we merge the metro and tramway data from the RATP with the light rail and tramway data from the SNCF. This aggregation gives what we call the “metro-rail-tramway” (MRT) layer, which we use as a single metro layer. We distinguish services that go in opposite directions or towards different branches, and we are able to identify them as a particular route because they share the same short name. We use the services that were available on Monday 26/5/2014, and we exclude any bus route that is completely subsumed by another route.

The RATP data provide extremely detailed transfer times between stops. We use this information to reconstruct connections between routes. With the exception of connections in the central station Châtelet-Les Halles, we ignore transfers that take longer than 8 minutes and consider the two corresponding nodes as different entities. By studying the cumulative distribution for the walking distances of these connections data, we find that the 99th percentile corresponds roughly to $d_w = 250m$. More precisely, $F(250m) \approx 0.986$. Motivated by this calculation, we allow a maximum walking distance of $d_w = 250m$ for the coarse-graining procedure that we need to construct the multilayer networks in New York City (NYC) and Tokyo.

For Paris, the transfer data allow a more accurate method than for NYC and Tokyo. We construct a transfer network in which (i) a transfer is defined between two stops and (ii) two stops are located in the same geographical position. We associate a single node to each connected component of the transfer network, and each isolated stop constitutes a single node. We weight the edges using travel times. Relatively large connection areas emerge from our choices. They correctly reflect possible choices that are suggested to travelers on the Paris metro map (32) (via the representation of intra-station walking paths as white edges).

Finally, we also couple nodes if (i) they are closer than $d_w = 250$ meters and if they come from RATP data and are labeled with the same name or (ii) they come from SNCF data (where transfer times between lines were not provided). Intra-route connections are possible if routes share the same node. As for NYC, the trip length is equal to the sum of the mean travel times

between consecutive stops.

New York City

We construct the NYC multilayer network with data from the Metropolitan Transportation Authority (MTA) Data Feeds (42), which provide a snapshot of their service for different days using the General Transit Feed Specification (GTFS) (43). We use the services that were available on Monday 1/12/2014. For the metro layer, we exclude six routes that represent shuttles or other special services.

For the bus layer, we integrate the data from the five NYC boroughs with multi-borough data of the subsidiary MTA Bus Company. We perform a coarse-graining procedure in which (i) we couple bus stops to metro locations if they lie within a walking distance of $d_w = 250$ meters and then (ii) combine bus stops into a single entity if their distance is less than d_w . We therefore allow intra-route connections if two routes share the same stop (where coupled stops count as a single stop). As with Paris, the trip length is given by the sum of the mean travel times between consecutive stops.

Tokyo

The Tokyo metro data set, which we have extracted from Wikipedia, allows us to study another of the world’s largest metropolitan networks (38). The bus data for Tokyo (for 2010) is provided freely by the Japanese Ministry of Land, Infrastructure, Transport, and Tourism (44). We use only Toei Bus lines (45), which serve central Tokyo.

This data set associates stops to bus routes, but it gives no information about the topology of the line. We reconstruct an approximate topology by generating a minimal spanning tree (46) among all stops of a route. We define the intra-route connections using the same coarse-graining procedure as for NYC. We again combine nodes that are within a walking distance of $d_w = 250$ meters from each other. We have no information on the travel times along the edges, so we estimate them from the trip length and using typical transportation speeds for the bus and metro in the Paris data. Because one can approximate each layers’ speed by a log-normal distribution, we use the log-average $\bar{v} = \exp \langle \log(v) \rangle$ as the typical speed associated to each mode of transport. The values that we find are $v_B \approx 14.0$ km/h for the bus and ($v_M \approx 23.4$ km/h) for the metro.

Definitions of Paths and Information Entropy

From the perspective of information processing, one can quantify the difficulty of navigating in an urban transportation network using a measure of “search information” S (16). This measure represents the amount of information that is needed for encoding a path from a route s to a route t that follows a simplest path. We define a *simplest path* as a shortest path in the dual space of a transportation network. In the present context, the dual space (which is also called

“P-space” (39)) is the network in which the nodes represent routes and the edges represent possible intersections (or connections between different lines) among those routes. There can be many degenerate simplest paths between two routes, and one can define the information entropy computed on these paths to be (16)

$$S^D(s, t) = -\log_2 \left[\sum_{\{p(s, t)\}} \left(\frac{1}{k_s} \prod_{n \in \{p(s, t)\}} \frac{1}{k_n - 1} \right) \right], \quad (4)$$

where the set $\{p(s, t)\}$ includes all $D(s, t)$ degenerate simplest paths between routes s and t . The quantity k_s indicates the total number of connections that emanate from route s . On a path, one has a choice between $k_n - 1$ routes (i.e., excluding the route whence the path emanates).

When one is seeking an optimal trajectory, the degeneracy is not necessarily a significant factor, and we focus on a trip from a node i to another node j in real space. Among all of the degenerate simplest paths, we pick a shortest one $p(s, t, i, j)$ and consider its entropy

$$S(i, s; j, t) = \log_2(k_s) + \sum_{n \in \{p(s, t, i, j)\}} \log_2(k_n - 1). \quad (5)$$

Averaging over all nodes $i \in s$ and $j \in t$ yields the main quantity that we used in the main text:

$$\bar{S}(s, t) = \langle S(s, t, i, j) \rangle. \quad (6)$$

In theory, it is also possible to weight this mean using real flows. Unfortunately, for this particular case, traditional data sources are insufficient, as they tend to describe the mobility of commuters, who do not necessarily rely on maps for their daily journeys. If we assume that all of the degenerate paths contribute equally to the entropy, then we obtain the following approximate relation between the entropies:

$$\bar{S}(s, t) \approx S^D(s, t) + \log_2[D(s, t)]. \quad (7)$$

Deviations from the (approximate) equality (7) indicate differences between the various degenerate paths (see Fig. 10).

From a map user’s perspective, the existence of several alternative simplest paths is not necessarily a significant factor, as one only needs a single simplest path for successful transportation from origin to destination. The natural choice is the fastest simplest path path, which is granted as unique for continuous travel-times. Consequently, we use the entropy in Eq. 1 rather than the one proposed in Ref. (16)

Acknowledgements

We thank M. Kivelä for helpful comments on an early draft of this paper. We thank P. Brodersen, A. Solé Ribalta and the anonymous reviewers for useful comments. RG thanks L. Alessandretti for suggesting the use of SNCF data. MB thanks Dr. Y. Shibata for helping to find and process the Tokyo data. All authors are supported by the European Commission FET-Proactive project PLEXMATH (Grant No. 317614). The authors declare no competing financial interests.

References

1. United Nations World Urbanization Prospects: www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html (accessed 10/01/2015).
2. R. Gallotti, M. Barthelemy, Anatomy and efficiency of urban multimodal mobility. *Sci. Rep.* **4**, 6911 (2014).
3. C. Roth, S. M. Kang, M. Batty, M. Barthelemy, A long-time limit for world subway networks. *J. R. Soc. Interface* **6**, 1–8 (2012).
4. M. J. Roberts, E. J. Newton, F. D. Lagattolla, S. Hughes, M. C. Hasler, Objective versus subjective measures of Paris metro map usability: Investigating traditional octolinear versus all-curves schematics. *Int. J. Hum.-Comput. St.* **71**, 363–386 (2013).
5. G. A. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. rev.* **63.2** 81 (1956).
6. R. I. M. Dunbar, Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **22**, 469–493 (1992).
7. J. Sarämäki, E. A. Leicht, E. López, S. G. B. Roberts, F. Reed-Tsochas, R. I. M. Dunbar, Persistence of social signatures in human communication. *Proc. Natl Acad. Sci. USA* **111**, 942–947 (2014).
8. R. I. M. Dunbar, Coevolution of neocortical size, group size and language in humans. *Behav. Brain. Sci.* **16**, 681–694 (1993).
9. B. Gonçalves, N. Perra, A. Vespignani, Modeling users' activity on Twitter networks: Validation of Dunbar's number. *PLoS ONE* **6**, e22656 (2011).
10. S. Srinivas, S. C. Hirtle, Knowledge based schematization of route directions. *Spatial Cognition V - Reasoning, Action, Interaction*. Springer Berlin Heidelberg, 346–364 (2007).
11. S. J. Luck, E. K. Vogel, The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
12. M. Burch, K. Kurzahls, M. Raschke, T. Blascheck, D. Weiskopf, How do people read metro maps? An eye tracking study. *Proceedings of First Workshop on Schematic Mapping* <https://sites.google.com/site/schematicmapping/Burch-EyeTracking.pdf> (2014).
13. M. Burch, K. Kurzahls, D. Weiskopf, Visual task solution strategies in public transport maps. *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research*, 32–36 (2014).

14. H. F. Credidio, E. N. Teixeira, S. D. S. Reis, A. Moreira, J. Andrade Jr, Statistical patterns of visual search for hidden objects. *Sci. Rep.* **2**, 920 (2012).
15. E. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, C. D. Frith, Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci. USA*, **97**, 4398–4403 (2000).
16. M. Rosvall, A. Trusina, P. Minnhagen, K. Sneppen, Networks and cities: An information perspective. *Phys. Rev. Lett.* **94**, 028701 (2005).
17. M. Wardman, Public transport values of time. *Transp. Policy* **11**, 363–377 (2004).
18. M. P. Viana, E. Strano, P. Bordin, M. Barthelemy, The simplicity of planar networks. *Sci. Rep.* **3**, 3495 (2013).
19. A. Baddeley, Working memory: Looking back and looking forward. *Nat. Rev. Neurosci.* **4**, 829–839 (2003).
20. F. Paas, J. E. Tuovinen, H. Tabbers, P. W. M. Van Gerven, Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* **38**, 63–71 (2003).
21. V. Latora, M. Marchiori, Is the Boston subway a small-world network? *Physica A* **314**, 109–113 (2001).
22. P. Angeloudis, D. Fisk, Large subway systems as complex networks. *Physica A* **367**, 553–558 (2006).
23. K. Lee, W.-S. Jung, J. S. Park, M. Y. Choi, Statistical analysis of the metropolitan Seoul subway system: Network structure and passenger flows. *Physica A* **387**, 6231–6234 (2008).
24. S. Derrible, C. Kennedy, The complexity and robustness of metro networks. *Physica A* **389**, 3678–3691 (2010).
25. M. P. Rombach, M. A. Porter, J. H. Fowler, P. J. Mucha, Core-periphery structure in networks. *SIAM J. App. Math.* **74**, 167–190 (2014).
26. R. Louf, C. Roth, M. Barthelemy, Scaling in transportation networks. *PLoS One* **9**, e102007 (2014).
27. S. H. Lee, P. Holme, Exploring maps with greedy navigators. *Phys. Rev. Lett.* **108**, 128701 (2012).
28. T. Hoffmann, R. Lambiotte, M. A. Porter, Decentralized routing on spatial networks with stochastic edge weights, *Phys. Rev. E* **88**, 022815 (2013).

29. M. De Domenico, A. Solé-Ribalta, S. Gomez, A. Arenas, Navigability of interconnected networks under random failures. *Proc. Natl. Acad. Sci. USA* **111**, 8351–8356 (2014).
30. A. Baddeley, The magical number seven: Still magic after all these years? *Psychological Review*, **101**, 353–356 (1994).
31. R. G. Morris, M. Barthelemy, Transport on coupled spatial networks. *Phys. Rev. Lett.* **109**, 128703 (2012).
32. RATP website:
http://www.ratp.fr/informer/pdf/orienter/f_plan.php (accessed on 14/1/2015).
33. M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer networks. *J. Complex Netw.* **2**, 203–271 (2014).
34. S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks. *Phys. Rep.* **544**, 1–122 (2014).
35. Wikimedia Commons:
http://commons.wikimedia.org/wiki/File:NYC_subway-4D.svg (accessed on 30/1/2015).
36. Transport For London: Bus spider maps:
https://tfl.gov.uk/maps_/bus-spider-maps (accessed on 4/6/2015).
37. M. De Domenico, A. Lima, M. C González, A. Arenas Personalized routing for multitudes in smart cities *EPJ Data Sci.*, **4**, (2015).
38. Wikipedia:
<http://www.wikipedia.org>
39. M. Barthelemy, Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
40. Data RATP:
<http://data.ratp.fr/fr/les-donnees.html> (accessed on 12/7/2014).
41. Data SNCF:
<http://files.transilien.com/horaires/gtfs/> (accessed on 1/10/2014).
42. MTA website:
<http://web.mta.info/> (accessed on 29/11/2014).
43. Google’s GTFS:
<https://developers.google.com/transit/gtfs/> (accessed on 15/1/2014).

44. Tokyo MLIT website:
<http://nlftp.mlit.go.jp/ksj-e/gml/datalist/KsjTmplt-P11.html> (accessed on 1/10/2014).
45. Toei line website:
<http://www.kotsu.metro.tokyo.jp/eng/> (accessed on 15/1/2014).
46. J. Clark, D.A. Holton, *A first look at graph theory (Vol. 1)*. Teaneck, NJ: World Scientific (1991).

1 Supplementary Materials

City	Abbreviation	Nodes	Edges	N	K_{tot}	P-diameter
New York City	NY	433	497	22	161	2
Paris	Pa	299	355	16	78	3
Tokyo	To	217	262	13	56	2
London	Lo	266	308	11	48	2
Madrid	Ma	209	240	12	38	2
Barcelona	Ba	139	165	11	37	2
Moscow	Mo	134	156	11	35	2
Seoul	Se	420	466	12	35	3
Shanghai	Sh	239	264	11	35	3
Mexico City	MC	147	164	11	31	2
Berlin	Ber	170	282	10	29	2
Chicago	Ch	167	222	8	25	2
Osaka	Os	108	123	9	24	2
Beijing	Bei	163	176	13	21	4
Hong Kong	HK	84	87	10	12	4

Table 1: **Network characteristics of the largest connected component for the 15 largest metropolitan systems in the world.** The number of routes N and connections K_{tot} , respectively, yield nodes and edges in the dual space. We list cities from most connections to fewest connections between different lines. The number of connections K_{tot} is the key quantity from the perspective of information processing. (See the right panel of Fig. 2.) *P-diameter* indicates the network diameter in dual space. It is equal to 2 for 10 of the 15 networks, and one additionally obtains a value of 2 in Paris if one cuts “3bis” (a four-stop line).

Network	Nodes	Edges	N	K_{tot}
NYC Metro	412	512	20	162
NYC Bus	5306	8435	309	6092
NYC multilayer	5332	8804	330	8461
Paris MRT	629	765	28	162
Paris Bus	3842	5567	277	2708
Paris multilayer	4037	6142	305	4292
Tokyo Metro	217	262	13	56
Tokyo Bus	1359	1663	153	1275
Tokyo multilayer	1422	1908	166	1831

Table 2: **Network characteristics of the Bus-Metro multilayer networks.** Similarly to Table 1, we show the number of routes N and connections K_{tot} , and the nodes and edges of the dual space. We note that there is a difference of an order of magnitude between the dimension of the Metro and the Bus layers, which represents a huge jump in complexity challenging people’s ability to navigate the multilayer transport networks.

	$C = 1$	$C = 2$	$C = 3$
NYC	85.2%	14.8%	0%
Paris MRT	43.0%	48.5%	8.5%
Tokyo	72.8%	28.2%	0%

Table 3: **Structure of Simplest Paths in Three Metro Systems.** We compare the number of connections in the simplest paths for the metro systems of the three megacities (New York City, Paris, and Tokyo) that we consider in detail. Only Paris has paths with more than 2 connections. A negligible fraction (not displayed) of paths have 4 or 5 connections.

	C=1	C=2
NYC	99.4%	0.5%
Paris	86.4%	13.6%
Tokyo	79.9%	20.1%

Table 4: **Length proportions of the paths with $\bar{S} < 8.1$ bits.** For the three megacities that we consider in this paper, about 20% of the trips have an information entropy that is lower than the threshold of 8.1 bits. Such trips predominantly have only a single connection. When there are more, the starting route has a limited number of connections (see Fig. 9).

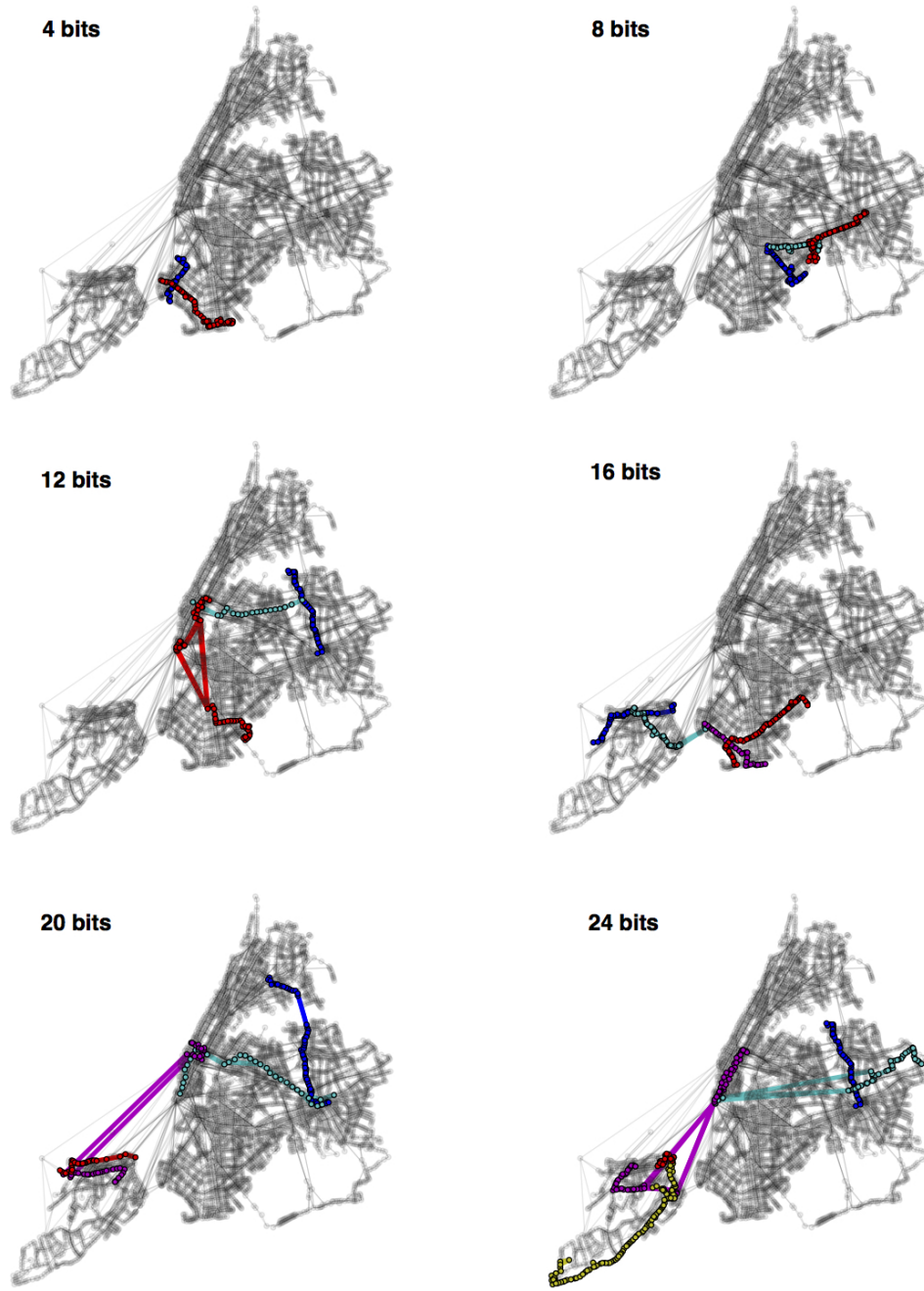


Figure 5: **Examples of paths with growing $\bar{S}(s, t)$.** For the New York City multilayer network, we show examples with increasing complexity: $\bar{S}(s, t)$ ranges from 4 bits to 24 bits. We color the starting bus line s in blue and the destination line t in red.

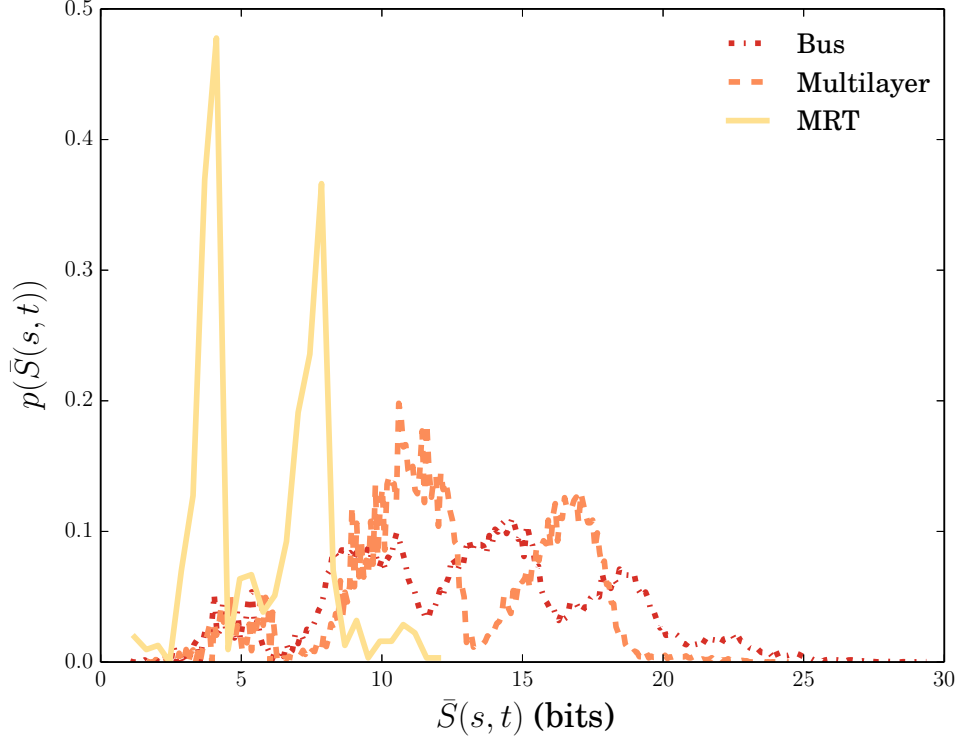


Figure 6: **Entropy distribution for MRT layer, bus layer, and complete multilayer network in Paris.** For the multilayer network, we restrict the distribution to trips whose origin and destination are each in the bus layer. We see that the effect of multiplexity on the bus layer is to shift the peaks to the right, and we also obtain larger peaks for smaller values of C (see Fig. 7).

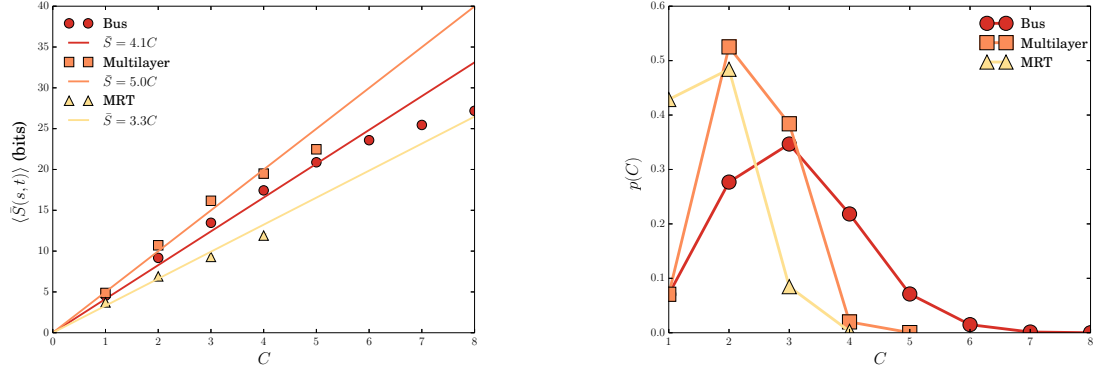


Figure 7: **Effects of multiplexity.** (Left) The values of \bar{S} grow with C . The growth is larger for the multilayer network (in the sense that it is characterized by the highest value of $\langle k \rangle$), smaller for the bus layer (where the lines have fewer connections), and is smallest for the MRT layer. (Right) Conversely, the mean path length is smaller for the bus monolayer network than for the multilayer network, in which the bus service interacts with the (longer-range) lines in the MRT layer.

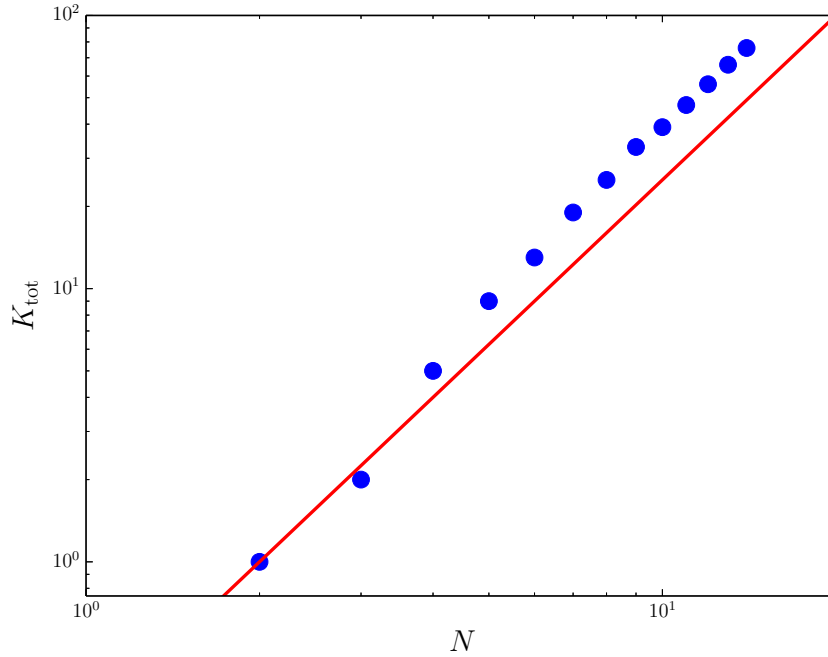


Figure 8: **Growth of the Paris Metro Network.** Letting the network grow with its historical progression from line 1 to line 14, we see that the number of connections in the dual space K_{tot} (blue dots) grows similarly to a lattice (red line), which would have $(N/2)^2$ intersections.

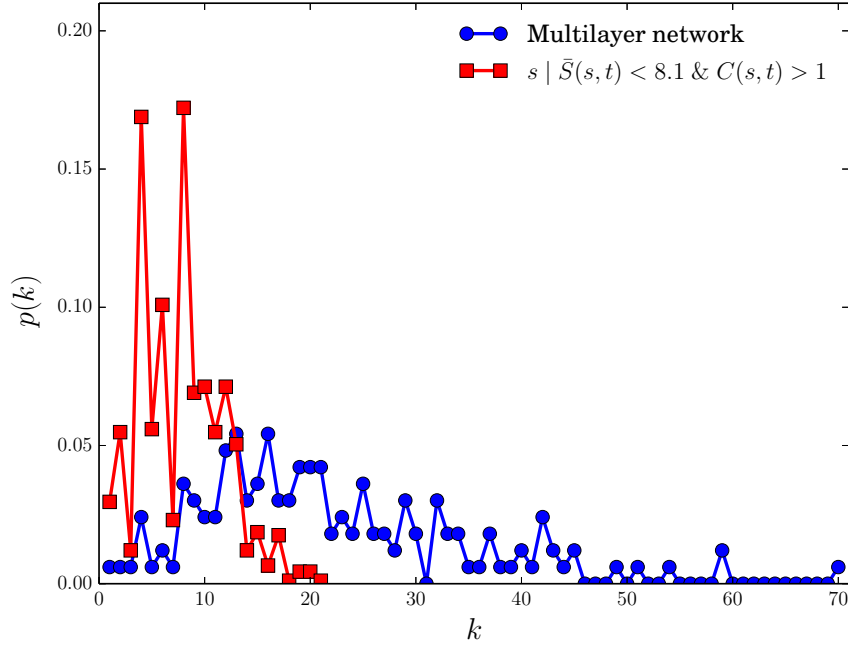


Figure 9: **Dual-space degree of low-information starting points for paths with $C = 2$ in the Tokyo multilayer network.** As one can see in Table 4, most of the trips below the cognitive limit have $C = 1$ connections. The trips below the cognitive threshold with $C > 1$ connections are characterized by a low connectivity of the origin route. We show this feature for the Tokyo case. We see that $C = 2$ for 20.1% of the trips that are below the threshold of 8.1 bits. For this fraction of trips $s \rightarrow t$, the degrees k in the dual space of origin routes s (red squares) are small in comparison to the degrees of all routes on the whole multilayer network (blue circles).

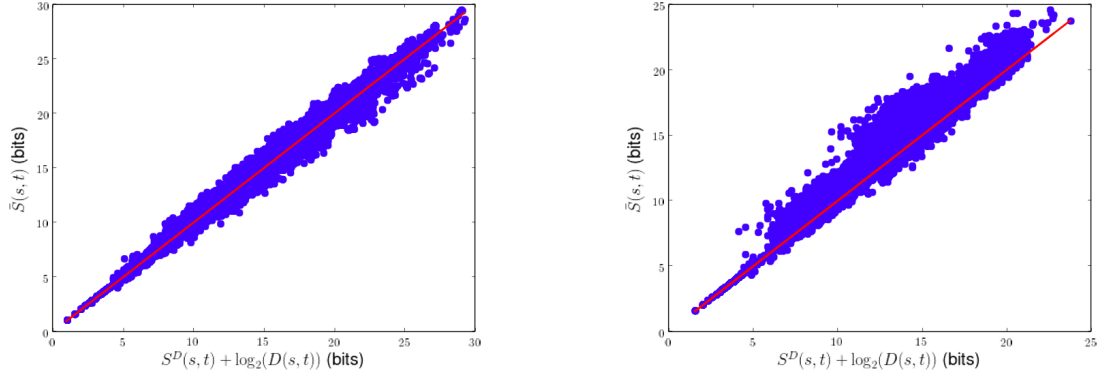


Figure 10: **Empirical validation of Eq. (7).** Comparing the (left) bus monolayer network to the (right) multilayer network in Paris, we note that including the metro-rail-tram (MRT) layer yields larger fluctuations. The mean square deviation is 0.34 bits for the bus layer and 0.82 bits of the multilayer network that contains both bus and MRT modes. This suggests that for the same route pair (s, t) , different paths become optimal for different origins i and destinations j .

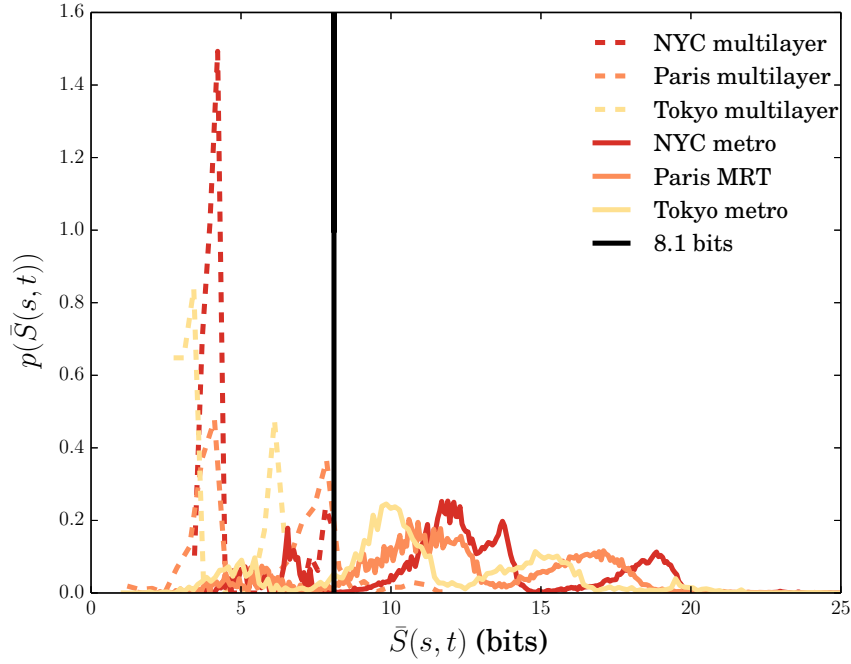


Figure 11: **Information entropy of multilayer networks.** This figure represents the probability density distributions of $\bar{S}(s, t)$. In Fig. 4 of the main text, we show the associated cumulative distributions. Similar to Fig. 4, we associate one layer to bus routes and another to metro lines. The solid curves are associated with multilayer networks that include a metro layer for New York City, Paris, and Tokyo. The dashed curves are associated to all possible paths in a metro layer. We observe that every distribution is characterized by a peak structure, and every peak is associated to a number C of connections.